The geometry of the generalized gamma manifold and an application to medical imaging

Sana Rebbah, Florence Nicol, Stéphane Puechmorel

Article

The geometry of the generalized gamma manifold and an application to medical imaging

Sana Rebbah [1], Florence Nicol [2] and Stéphane Puechmorel [3,*], for the Alzheimer's Disease Neuroimaging Initiative*

[1] ENAC and INSERM ToNIC, Université de Toulouse; sana.rebbah@inserm.fr
[2] ENAC, Université de Toulouse, florence.nicol@enac.fr
[3] ENAC, Université de Toulouse, stephane.puechmorel@enac.fr
* Correspondence: stephane.puechmorel@enac.fr; Tel.: +33-5-62259503

Version June 18, 2019 submitted to Mathematics

Abstract: The Fisher information metric provides parameterized probability densities with a Riemannian manifold structure, yielding the so-called information geometry. The information geometry of the gamma manifold associated to the family of gamma distributions has been well studied. However, only a few results are known for the generalized gamma family, that adds an extra shape parameter. The present article gives some new results about the generalized gamma manifold. This paper also introduces an application in medical imaging that is the classification of Alzheimer's disease population. In the medical field, over the past two decades, a growing number of quantitative image analysis techniques have been developed, including histogram analysis, which is widely used to quantify the diffuse pathological changes of some neurological diseases. This method presents several drawbacks. Indeed, all the information included in the histogram is not used and the histogram is an overly simplistic estimate of a probability distribution. Thus, in this study we present how using information geometry and the generalized gamma manifold improved the performance of the classification of Alzheimer's disease population.

1. Introduction

The generalized gamma distribution was introduced in [1], and can be viewed as a special case of the Amoroso distribution [2] in which the location parameter is dropped [3]. Apart from the gamma distribution, it generalizes also the Weibull distribution and is of common use in survival models. The purpose of the present work is to investigate some information geometric properties of the generalized gamma family, especially when restricted to the gamma submanifold. First, in Section 2, the Fisher information as a Riemannian metric and results in the case of the gamma manifold will be briefly introduced. Next, the case of the generalized gamma manifold will be detailed, using an approach based on diffeomorphism groups. In section 4, the extrinsic curvature of the gamma submanifold will be computed. Finally, an example of application in the medical imaging domain will be given in the last section.

2. Information geometry and the gamma manifold

Information geometry deals with parameterized families of distributions whose parameters are understood as local coordinates on a manifold and provided with a Riemannian structure by the

*Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Fisher metric. In the sequel, Θ will be a smooth manifold and $(p_\theta), \theta \in \Theta$ a family of probability density functions defined on a common event space Ω and depending smoothly on the parameters θ. Thorough the paper, the Einstein summation convention on repeated indices will be used.

Definition 1. *The Fisher information metric on Θ is defined at point $\theta \in \Theta$ by the symmetric order 2 tensor:*

$$g = g_{ij} d\theta^i \otimes d\theta^j$$

where:

$$g_{ij} = E_{p_\theta}\left[\partial_{\theta_i} l \partial_{\theta_j} l\right], \, l(\theta) = \log p_\theta$$

When the support of the density functions p_θ does not depend on θ, the information metric can be rewritten as:

$$g_{ij} = -E_{p_\theta}\left[\partial_{\theta_i}\partial_{\theta_j} l\right] \tag{1}$$

It gives rise to a Riemannian metric on Θ.

When the underlying event space Ω is also a smooth manifold, the Fisher metric has a classical nice invariance property, that corresponds to information preservation by sufficient statistics:

Proposition 1. *Let $\tilde{\Omega}$ be a smooth manifold and $\Phi\colon \Omega \to \tilde{\Omega}$ be a smooth diffeomorphism. Let $\tilde{g}$ be the Fisher information metric associated to the image family $\Phi_* p_\theta$ defined on the event space $\tilde{\Omega}$. Then $\tilde{g} = g$.*

The Fisher metric has a very simple expression when the parameterized family p_θ is of natural exponential type. In such a case, assuming for the sake of simplicity that Θ and Ω are open subsets of finite dimensional real vector spaces, the density function p_θ can be written as:

$$p_\theta(x) = \exp(\langle \theta, F(x) \rangle - \phi(\theta) + g(x)) \tag{2}$$

The function ϕ in eq. 2 is called the potential function of the density and an immediate application of the definition (1) yields for the expression of the Fisher information metric:

$$g_{ij}(\theta) = \frac{\partial^2 \phi}{\partial_{\theta_i} \partial_{\theta_j}}(\theta) \tag{3}$$

A manifold with such a Riemannian metric is referred to as a Hessian structure [4]. Many important tools from Riemannian geometry, like the Levi-Civita connection, are greatly simplified within this frame. In the sequel, all partial derivatives ∂_{θ_i} will be abbreviated by ∂_i.

Proposition 2. *For a parameterized density family $p_\theta, \theta \in \Theta$ pertaining to the natural exponential class with potential function Ψ, the Christoffel symbols of the first kind of the associated Hessian structure are given by [5]:*

$$\Gamma_{ijk} = \frac{1}{2} \partial_i \partial_j \partial_k \phi$$

The gamma distribution can be written as a natural exponential family on two parameters (α, λ), defined on the parameter space $\mathbb{R}^{+*} \times \mathbb{R}^{+*}$ by:

Definition 2. *The gamma distribution is the probability law on $\mathbb{R}^{+*}$ with density relative to the Lebesgue measure given by:*

$$p(x; \alpha, \lambda) = \frac{1}{\Gamma(\lambda)\alpha^\lambda} x^{\lambda - 1} e^{-\frac{x}{\alpha}}, \, x > 0 \tag{4}$$

with parameters $\alpha > 0, \lambda > 0$.

The next proposition comes directly from the definition:

Proposition 3. *The gamma distribution defines a natural exponential family with natural parameters λ and $\eta = \alpha^{-1}$ and potential function $\phi(\eta, \lambda) = \log(\Gamma(\lambda)) - \lambda \log(\eta)$.*

Using 3, the Fisher metric is obtained by a straightforward computation:

$$g(\eta, \lambda) = \begin{pmatrix} \frac{\lambda}{\eta^2} & -\frac{1}{\eta} \\ -\frac{1}{\eta} & \psi'(\lambda) \end{pmatrix} \tag{5}$$

where ψ is the digamma function.

It is sometimes convenient to perform a change of parameterization in order to have a diagonal form for the metric. The next proposition is of common use and allows the computation of a pullback metric in local coordinates:

Proposition 4. *Let $\mathcal{M}$ be a smooth manifold and $(\mathcal{N}, g)$ be a smooth Riemannian manifold. For a smooth diffeomorphism $f\colon \mathcal{M} \to \mathcal{N}$, the pullback metric f^*g has matrix expressed in local coordinates at the point $m \in \mathcal{M}$ by:*

$$J_f^t(m) G(f(m)) J_f(m) \tag{6}$$

with $J_f(m)$ the jacobian matrix of f at m and $G(n)$ the matrix of the metric g at $n \in \mathcal{N}$.

Performing the change of parameterization: $f\colon (\mu, \beta) \mapsto (\eta = \beta/\mu, \lambda = \beta)$ yields:

$$J_f(\mu, \beta) = \begin{pmatrix} -\frac{\mu}{\beta^2} & \frac{1}{\beta} \\ 0 & 1 \end{pmatrix}$$

Using prop. 4 then gives for the pullback metric matrix:

$$G(\mu, \beta) = \begin{pmatrix} \frac{\beta}{\mu^2} & 0 \\ 0 & \psi(\beta)' - \frac{1}{\beta} \end{pmatrix}.$$

The information geometry of the gamma distribution is studied in details in [6], with explicit calculations of the Christoffel symbols and the geodesic equation.

3. The geometry of the generalized gamma manifold

While the gamma distribution is well suited to study departure to full randomness has pointed out in [6], it is not general enough in many applications. In particular, the Weibull distribution, that also generalizes the exponential distribution is not a gamma distribution. A more general family was thus introduced, by adding a power term.

Definition 3. *The generalized gamma distribution is the probability measure on $\mathbb{R}^+$ with density respective to the Lesbesgue measure given by:*

$$p(x; \alpha, \lambda, \beta) = \frac{\beta x^{\beta\lambda - 1}}{\alpha^{\beta\lambda}\Gamma(\lambda)} e^{-\left(\frac{x}{\alpha}\right)^{\beta}}, \; x > 0 \tag{7}$$

where $\alpha > 0, \lambda > 0, \beta > 0$.

Due to the exponent β, the generalized gamma distribution does not define a natural exponential family. However, letting β fixed, the mapping $\Phi_\beta\colon x \mapsto x^\beta$ is a diffeomorphism of $\mathbb{R}^+$ to itself, and the image density of $p(\alpha, \lambda, \beta)$ under Φ_β is a gamma density with parameters (α^β, λ). For any $\kappa > 0$, the submanifold $\beta = \kappa$ of the generalized gamma manifold is diffeomorphic to the gamma manifold. Using the invariance of the Fisher metric under diffeomorphisms, the induced metric on the above submanifold can be obtained.

Proposition 5. *Let $\kappa > 0$ be a fixed real number. The induced Fisher metric G_κ on the submanifold $(\alpha, \lambda, \kappa)$ of the generalized gamma manifold is given in local coordinates by:*

$$G_\kappa(\alpha, \lambda) = \begin{pmatrix} \frac{\lambda\kappa^2}{\alpha^2} & -\frac{\kappa}{\alpha} \\ -\frac{\kappa}{\alpha} & \psi'(\lambda) \end{pmatrix}.$$

Proof. In local coordinates (α^κ, λ), the Fisher metric of a gamma distribution manifold (α^κ, λ) is

$$G_\kappa(\alpha^\kappa, \lambda) = \begin{pmatrix} \frac{\lambda}{\alpha^{2\kappa}} & -\frac{1}{\alpha^\kappa} \\ -\frac{1}{\alpha^\kappa} & \psi'(\lambda) \end{pmatrix}.$$

The Jacobian matrix of the transformation $(\alpha, \lambda) \to (\alpha^\kappa, \lambda)$ is the matrix $J = \operatorname{diag}(\kappa\alpha^{\kappa-1}, 1)$ and the change of parametrization yields:

$$G_\kappa(\alpha, \lambda) = J^t G_\kappa(\alpha^\kappa, \lambda) J.$$

The Fisher metric on the submanifold $(\alpha, \lambda, \kappa)$ is directly obtained from the invariance by using the diffeomorphism $\Phi_\beta \colon x \mapsto x^\beta$. $\square$

Proposition 6. *In local coordinates, the fisher information metric of the generalized gamma manifold is given by:*

$$G(\alpha, \lambda, \beta) = \begin{pmatrix} \frac{\beta^2\lambda}{\alpha^2} & -\frac{\beta}{\alpha} & \frac{-\lambda\psi(\lambda)-1}{\alpha} \\ -\frac{\beta}{\alpha} & \psi'(\lambda) & -\frac{\psi(\lambda)}{\beta} \\ \frac{-\lambda\psi(\lambda)-1}{\alpha} & -\frac{\psi(\lambda)}{\beta} & \frac{\lambda\psi(\lambda)^2+2\psi(\lambda)+\lambda\psi'(\lambda)+1}{\beta^2} \end{pmatrix} \tag{8}$$

Proof. The 2×2 submatrix corresponding to local coordinates α, λ has already been obtained in prop. 5. The remaining terms can be computed by differentiating the log likelihood function twice, but an alternative will be given below in a more general setting. $\square$

The usual definition of the generalized gamma distribution 3 does stems from the gamma one by a simple change of variable, thus making some computation less natural. Starting with the above diffeomorphism Φ_β and applying it to a gamma distribution yields an equivalent, but more intuitive form. Furthermore, it is advisable to express the gamma density as a natural exponential family distribution:

$$p(x; \eta, \lambda) = \frac{\eta^\lambda\, x^{\lambda-1}\, e^{-\eta x}}{\Gamma(\lambda)}, \quad x > 0,$$

where $\lambda > 0$, $\eta > 0$ are the natural parameters of the distribution.

Definition 4. *The generalized gamma distribution on $\mathbb{R}^+$ is the probability measure with density:*

$$p(x; \eta, \lambda, \beta) = \frac{\beta\, \eta^\lambda\, x^{\beta\lambda-1}\, e^{-\eta x^\beta}}{\Gamma(\lambda)}, \quad x > 0,$$

with $\eta > 0$, $\lambda > 0$ and $\beta > 0$.

Due the the invariance by diffeomorphism property of the Fisher information metric, the induced metric on the submanifolds $\beta = \text{cte}$ is independent of β, and is exactly the one of the gamma manifold, here given by:

$$g(\eta, \lambda) = \begin{pmatrix} \frac{\lambda}{\eta^2} & -\frac{1}{\eta} \\ -\frac{1}{\eta} & \psi'(\lambda) \end{pmatrix}. \tag{9}$$

An important fact about the family of diffeomorphisms Φ_β is the group property $\Phi_{\beta_1} \circ \Phi_{\beta_2} = \Phi_{\beta_1 \beta_2}$. It turns out that all the computation can be conducted in a general Lie group setting, as detailed below. Let $p_\theta, \theta \in \Theta$, be a parameterized family of probability densities defined on an open subset U of $\mathbb{R}^n$ and let G be a Lie group action on U by diffeomorphisms preserving orientation. For any g in G and θ in Θ, the image density $\tilde{p}_{g,\theta}$ under the diffeomorphism $x \in U \mapsto \xi(g, x) = g.x$ is given by:

$$\forall x \in U, \tilde{p}_{g,\theta}(x) = p_\theta(\xi(g, x)) \left|\partial_2 \xi(g, x)\right|.$$

Note that, in this paper, we consider increasing monotone diffeomorphisms. For simplicity of calculus, the absolute value may be remove in the above expression. Denoting $\tilde{l}(x, \theta, g)$ the log-likelihood of $\tilde{p}_{g,\theta}(x)$ and $l(x, \theta)$ the one of $p_\theta(x)$, it comes, by obvious computation:

$$\forall x \in U, \tilde{l}(x, \theta, g) = l(\xi(g, x), \theta) + \log \partial_2 \xi(g, x).$$

Throughout the document, the symbol ∂_i stands for the partial derivative with respect to the i-th variable. Higher order derivatives are written similarly as $\partial_{i \ldots i, j \ldots j, \ldots}$ by repeating the variable k times to indicate a partial derivative of order k.

Proposition 7. *For any $x \in U$, $g \in G$:*

$$\partial_1 \xi(g, x) = \partial_1 \xi(e, \xi(g, x)) T_g R_{g^{-1}}$$

where e is the identity of G and R_g is the right translation mapping $h \in G \mapsto R_g.h = h.g$.

Proof. Since ξ comes from a group action:

$$\xi(h, \xi(g, x)) = \xi(h.g, x).$$

Then, taking the derivative with respect to h at identity:

$$\partial_1 \xi(e, \xi(g, x)) = \partial_1 \xi(g, x) T_e R_g.$$

Since $T_e R_g T_g R_{g^{-1}} = Id$ by the chain rule, the claimed result is proved. □

This property allows to compute the Fisher information metric in a convenient way.

Proposition 8. *The element $G_{g,\theta}$ of the Fisher metric of $\tilde{p}_{g,\theta}$ is given by:*

$$-\int_U \partial_{12} l(x, \theta) \partial_1 \xi(e, x) p_\theta(x) dx \; T_g R_{g^{-1}},$$

Proof. Since:

$$\tilde{l}(x, \theta, g) = l(\xi(g, x), \theta) + \log \partial_2 \xi(g, x),$$

it comes:

$$\partial_2 \tilde{l}(x, \theta, g) = \partial_2 l(\xi(g, x), \theta)$$

and thus:

$$\partial_{23} \tilde{l}(x, \theta, g) = \partial_{12} l(\xi(g, x), \theta) \partial_1 \xi(g, x).$$

Now, using prop. 7:
$$\partial_{23}\tilde{l}(x,\theta,g) = \partial_{12}l_\theta(\xi(g,x),\theta)\partial_1\xi(e,\xi(g,x))T_gR_{g^{-1}}.$$
Taking the expectation with respect to $\tilde{p}_{g,\theta}$ yields:
$$E[\partial_{23}] = \int_U \partial_{12}l(\xi(g,x),\theta)\partial_1\xi(e,\xi(g,x))\tilde{p}_{g,\theta}(x)dx\, T_gR_{g^{-1}}$$
and the result follows by the change of variable $y = \xi(g,x)$. □

The case of the elements $G_{g,g}$ is a little bit more complex, due to the non vanishing extra term in the log-likelihood $\tilde{l}(x,\theta,g)$. Taking the first derivative with respect to g yields:
$$\forall x \in U, \partial_3\tilde{l}(x,\theta,g) = \partial_1 l(\xi(g,x),\theta)\partial_1\xi(g,x) + \frac{\partial_{12}\xi(g,x)}{\partial_2\xi(g,x)}.$$
The second term in the right hand side can be further simplified using the next proposition, that is a direct consequence of prop. 7.

Proposition 9. *For any $\theta \in \Theta, g \in G, x \in U$:*
$$\partial_{12}\xi(e,\xi(g,x))\partial_2\xi(g,x) = \partial_{12}\xi(g,x)\, T_eR_g.$$

Applying it to the log-likelihood derivative and using again 7 yields:
$$\forall x \in U, \quad \partial_3\tilde{l}(x,\theta,g) = (\partial_1 l(\xi(g,x),\theta)\partial_1\xi(e,\xi(g,x)) + \partial_{12}\xi(e,\xi(g,x)))\, T_gR_{g^{-1}}.$$

Proposition 10. *The element $G_{g,g}$ of the Fisher metric of $\tilde{p}_{g,\theta}$ is given in matrix form by:*
$$T_gR^T_{g^{-1}} \int_U h_{g,\theta}(x)^T h_{g,\theta}(x)p_\theta(x)dx\, T_gR_{g^{-1}}$$
with:
$$h_{g,\theta}(x) = \partial_1 l(x,\theta)\partial_1\xi(e,x) + \partial_{12}\xi(e,x).$$

Proof. Starting with the definition:
$$G_{g,g} = E[(\partial_3\tilde{l})^T\,(\partial_3\tilde{l})]$$
the result follows after the change of variable $y = \xi(g,x)$ in the expectation. □

An important corollary of 8 and 10 is that the Fisher metric is right invariant with respect to the group action.

Propositions 8 and 10 allow to compute the coefficients $g_{\eta\beta}, g_{\lambda\beta}, g_{\beta\beta}$ in the Fisher metric, thus yielding the next proposition.

Proposition 11. *The Fisher information matrix in natural coordinates has coefficients:*

$$
\begin{aligned}
g_{\eta\eta} &= \frac{\lambda}{\eta^2} \\
g_{\eta\lambda} &= -\frac{1}{\eta} \\
g_{\lambda\lambda} &= \psi'(\lambda) \\
g_{\eta\beta} &= \frac{\lambda}{\eta\beta}\left(\psi(\lambda+1) - \log\eta\right) \\
g_{\lambda\beta} &= \frac{1}{\beta}\left(\log\eta - \psi(\lambda)\right) \\
g_{\beta\beta} &= \frac{1}{\beta^2}\left[1 + \lambda\log^2\eta - 2\lambda\psi(\lambda+1)\log\eta + \lambda\psi^2(\lambda+1) + \lambda\psi'(\lambda+1)\right]
\end{aligned}
$$

Recalling that the Christoffel symbols of the first kind for the Levi-Civita connection are obtained using the formula:

$$\Gamma_{kij} = \frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}\right)$$

one can obtain them as:

$$
\begin{array}{ccc}
\Gamma_{111} = -\frac{\lambda}{\eta^3} & \Gamma_{211} = \frac{1}{2\eta^2} & \Gamma_{311} = \frac{\lambda(-1+\log\eta-\psi(\lambda+1))}{\eta^2\beta} \\
\Gamma_{121} = \Gamma_{211} = \Gamma_{112} & \Gamma_{221} = \Gamma_{212} = 0 & \Gamma_{321} = \Gamma_{312} = \frac{1-\log\eta+\psi(\lambda+1)+\lambda\psi'(\lambda+1)}{2\eta\beta} \\
\Gamma_{122} = 0 & \Gamma_{222} = \frac{1}{2}\psi''(\lambda) & \Gamma_{322} = -\frac{\psi'(\lambda)}{\beta}
\end{array}
\tag{10}
$$

$$
\begin{gathered}
\Gamma_{131} = \Gamma_{113} = 0 \\
\Gamma_{231} = \Gamma_{213} = \frac{1+\log\eta-\psi(\lambda+1)-\lambda\psi'(\lambda+1)}{2\eta\beta} \\
\Gamma_{331} = \Gamma_{313} = \frac{\lambda(\log\eta-\psi(\lambda+1))}{\eta\beta^2} \\
\Gamma_{132} = \Gamma_{123} = \frac{-1-\log\eta+\psi(\lambda+1)+\lambda\psi'(\lambda+1)}{2\eta\beta} \\
\Gamma_{232} = \Gamma_{223} = 0 \\
\Gamma_{332} = \Gamma_{323} = \frac{\psi'(\lambda+1)(1-2\lambda\log\eta)-2\psi(\lambda+1)(\log\eta-\lambda\psi'(\lambda+1))+\log^2\eta+\psi(\lambda+1)^2+\lambda\psi''(\lambda+1)}{2\beta^2} \\
\Gamma_{133} = 0 \\
\Gamma_{233} = -\frac{-2\psi(\lambda+1)(\log\eta-\lambda\psi'(\lambda+1))+\psi'(\lambda+1)(1-2\lambda\log\eta)+\log\eta(\log\eta+2)+\psi(\lambda+1)^2-2\psi(\lambda)+\lambda\psi''(\lambda+1)}{2\beta^2} \\
\Gamma_{333} = -\frac{\lambda\log^2\eta+\lambda\left(-2\log\eta\psi(\lambda+1)+\psi(\lambda+1)^2+\psi'(\lambda+1)\right)+1}{\beta^3}
\end{gathered}
\tag{11}
$$

4. The gamma submanifold

The submanifolds $\beta = \text{cte}$ of the generalized gamma manifold are all isometric to the gamma manifold. This section is dedicated to the study of their properties using the Gauss-Codazzi equations. In the sequel, the generalized gamma manifold will be denoted by M while $N_\kappa, \kappa > 0$ will stand for the embedded submanifold $\beta = \kappa$.

Proposition 12. *The normal bundle to N_κ is generated at (η, λ) on the gamma submanifold by the vector:*

$$n(\eta,\lambda) = \left(-\eta(\lambda\psi'(\lambda)(\psi(\lambda+1) - \log(\eta)) + \log(\eta) - \psi(\lambda)), -1, \kappa\left(\lambda\psi'(\lambda) - 1\right)\right)$$

Proof. The matrix of the Fisher metric at (η, λ, β) can be written in block form as:

$$G(\eta,\lambda,\beta) = \begin{pmatrix} g(\eta,\lambda) & v \\ v^t & g_{\beta\beta} \end{pmatrix}$$

with:

$$g(\eta,\lambda)=\begin{pmatrix}\frac{\lambda}{\eta^2} & -\frac{1}{\eta}\\ -\frac{1}{\eta} & \psi'(\lambda)\end{pmatrix}$$

and

$$v=\begin{pmatrix}\frac{\lambda}{\eta\beta}(\psi(\lambda+1)-\log\eta)\\ \frac{1}{\beta}(\log\eta-\psi(\lambda))\end{pmatrix}$$

Any multiple of the vector:

$$(-g(\eta,\lambda)^{-1}v,1)$$

90 is normal to the tangent space to the submanifold N_κ. The result follows by simple computation. □

Let ∇ be the levi-civita connection of the gamma manifold and $\overline{\nabla}$ that of the generalized gamma. It is well known [7] (pp 60-63) that these two connections are related by the Gauss formula:

$$\forall X,Y\in TN_\kappa,\ \overline{\nabla}_X Y=\nabla_X Y+\mathcal{B}(X,Y) \tag{12}$$

where $\mathcal{B}$ is a symmetric bilinear form with values in the normal bundle. Letting $n=n^i e_i$ with $e_1=\partial_\eta, e_2=\partial_\lambda, e_3=\partial_\beta$, it comes, with $i,j=1\dots 2$:

$$g\left(\overline{\nabla}_{e_i}e_j,n\right)=n^k\overline{\Gamma}_{kij}=g\left(\nabla_{e_i}e_j,n\right)+g\left(B(e_i,e_j),n\right). \tag{13}$$

Since B takes its values in the normal bundle, it exists a smooth real value mapping $a_{ij}, i,j=1\dots 2$ such that $B(e_i,e_j)=a_{ij}n$ The equation 13 yields:

$$a_{ij}=\frac{n^k\overline{\Gamma}_{kij}}{g(n,n)}. \tag{14}$$

From [7] (p 63), the sectional curvature $\overline{K}(e_1,e_2)$ of M can be obtained from the one $K(e_1,e_2)$ of N_κ as:

$$\overline{K}(e_1,e_2)=K(e_1,e_2)+\frac{g\left(B(e_1,e_2),B(e_1,e_2)\right)-g\left(B(e_1,e_1),B(e_2,e_2)\right)}{g(e_1,e_1)g(e_2,e_2)-g(e_1,e_2)^2} \tag{15}$$

or:

$$\overline{K}(e_1,e_2)=K(e_1,e_2)+g(n,n)\frac{a_{12}^2-a_{11}a_{22}}{g_{11}g_{22}-g_{12}^2}. \tag{16}$$

Using the expressions if the Christoffel symbols and the metric, the coefficients a_{11},a_{12},a_{22} can be computed as:

$$a_{11}=\frac{2\lambda^(1-\lambda\psi'(\lambda))+1}{2\eta^2 D} \tag{17}$$

$$a_{12}=\frac{\lambda^2\psi'(\lambda)^2-\psi'(\lambda)-1}{2\eta D} \tag{18}$$

$$a_{22}=\frac{\psi'(\lambda)(1-\lambda\psi'(\lambda))-\psi''(\lambda)/2}{D} \tag{19}$$

with:

$$D=g(n,n)=(\lambda\psi'(\lambda)-1)(\psi'(\lambda)(\lambda^2\psi'(\lambda)-1)-1).$$

Finally:

$$g(n,n)\frac{a_{12}^2-a_{11}a_{22}}{g_{11}g_{22}-g_{12}^2}=F(\lambda)/G(\lambda) \tag{20}$$

with:

$$F(\lambda) = \lambda^4\psi'(\lambda)^4 - 2\lambda^2(2\lambda+1)\psi'(\lambda)^3 + \left(6\lambda^2+2\lambda+1\right)\psi'(\lambda)^2 \\ - 2\lambda(\lambda\psi''(\lambda)+2)\psi'(\lambda) + (2\lambda+1)\psi''(\lambda)+1$$

and:

$$G(\lambda) = 4(\lambda\psi'(\lambda)-1)^2\left(\psi'(\lambda)\left(\lambda^2\psi'(\lambda)-1\right)-1\right).$$

91 **Proposition 13.** *The term $a_{12}^2 - a_{11}a_{22}$ is strictly positive.*

Proof. Using the expressions of the coefficients:

$$a_{12}^2 - a_{11}a_{22} = \frac{1}{4\eta^2 D^2}\left(A(\lambda) + B(\lambda)C(\lambda)\right)$$

with:

$$\begin{aligned} A(\lambda) &= (\lambda^2\psi'(\lambda)^2 - \psi'(\lambda) - 1)^2 \\ B(\lambda) &= 2\lambda(1-\lambda\psi'(\lambda)) + 1 \\ C(\lambda) &= 2\psi'(\lambda)(-1+\lambda\phi'(\lambda)) + \psi''(\lambda). \end{aligned}$$

The ψ' function satisfies the next inequality [8]:

$$\frac{1}{\lambda} + \frac{1}{2\lambda^2} < \psi'(\lambda) < \frac{1}{\lambda} + \frac{1}{\lambda^2}$$

from which it comes:

$$-\frac{1}{2\lambda} > 1 - \lambda\psi'(\lambda) > -\frac{1}{\lambda}$$

and in turn:

$$0 > B(\lambda) > -1.$$

To obtain the sign of $C(\lambda)$, a different bound is needed for the polygamma function. Again from [8]:

$$\frac{(k-1)!}{(x+1)^k} + \frac{k!}{x^{k+1}} < \left|\psi^{(k)}\right| < \frac{(k-1)!}{(x+1/2)^k} + \frac{k!}{x^{k+1}}, \qquad k \geq 1. \tag{21}$$

Using the inequality 21, it comes:

$$\frac{\lambda+1}{\lambda(2\lambda+1)} < \lambda\psi'(\lambda) - 1$$

so that:

$$\left(\frac{1}{\lambda+1/2} + \frac{1}{\lambda^2}\right)\left(\frac{\lambda+1}{\lambda(2\lambda+1)}\right) < \psi'(\lambda)(-1+\lambda\phi'(\lambda)).$$

Using again 21 with $k=2$ yields finally:

$$C(\lambda) < -\frac{2}{\lambda^2(1+2\lambda)^2}.$$

92 Since both $B(\lambda)$ and $C(\lambda)$ are strictly negative, $A(\lambda) + B(\lambda)C(\lambda)$ is strictly positive as claimed. □

Proposition 14. *The sectional curvature of the generalized gamma manifold in the (e_1, e_2) satisfies:*

$$\overline{K}(e_1, e_2) \underset{\lambda->0^+}{\rightarrow} \frac{12-\pi^2}{2(\pi^2-6)}.$$

Proof. The sectional curvature of the gamma manifold satisfies [6]:

$$K(e_1, e_2) \underset{\lambda->0^+}{\rightarrow} -\frac{1}{2}.$$

It is thus only needed to estimate the limit of (20) when $\lambda \to 0^+$. The asymptotics of the polygamma functions at 0 are given by:

$$\psi'(\lambda) = \frac{1}{\lambda^2} + \psi'(1) + o(1),$$
$$\psi''(\lambda) = -\frac{2}{\lambda^3} + \psi''(1) + o(1).$$

The term:

$$F(\lambda) = \lambda^4 \psi'(\lambda)^4 - 2\lambda^2(2\lambda+1)\psi'(\lambda)^3 + \left(6\lambda^2 + 2\lambda + 1\right)\psi'(\lambda)^2 - 2\lambda(\lambda\psi''(\lambda)+2)\psi'(\lambda) + (2\lambda+1)\psi''(\lambda) + 1$$

can thus be approximated by:

$$\left(\pi^8 x^6 - 24\pi^6 x^5 + 12\pi^6 x^4 + 216\pi^4 x^4 - 432\pi^2 x^4 \psi''(1) - 360\pi^4 x^3 - 864\pi^2 x^3 + 2592 x^3 \psi''(1) + 36\pi^4 x^2 + 2592\pi^2 x^2 + 1296 x^2 - 1296 x^2 \psi''(1) - 864\pi^2 x - 5184 x + 2592\right) / (1296 x^2)$$

and the term :

$$G(\lambda) = 4(\lambda\psi'(\lambda) - 1)^2 \left(\psi'(\lambda)\left(\lambda^2\psi'(\lambda) - 1\right) - 1\right).$$

is approximated by:

$$\frac{\left(\pi^2 x^2 - 6x + 6\right)^2 \left(\pi^4 x^2 + 6\pi^2 - 36\right)}{324 x^2}$$

Finally, the quotient $F(\lambda)/G(\lambda)$ is equal at $\lambda = 0$ to

$$\frac{3}{\pi^2 - 6}$$

and the result follows by summation with $-1/2$. □

It is conjectured that the sectional curvature of the generalized gamma manifold in the directions $\partial_\eta, \partial_\lambda$ is strictly positive, bounded from above by $1/2$ as it appears to be the case numerically.

5. Medical imaging application

Magnetic Resonance Imaging (MRI) seeks to identify, localize and measure different parts of the anatomy of the central nervous system, and has been demonstrated as a valid marker of neurodegenerative diseases such as Alzheimer's disease, the most common cause of dementia [9–11]. Indeed, brain atrophy measured by structural MRI has been proposed as a surrogate marker for the early diagnosis of Alzheimer's disease [12,13].

Many of these studies limited their work by using central tendency measures such as the mean or the median and more recent ones used histogram-analysis [14,15] in order to represent a biomarker rather than using the biomarker probability distribution of the whole brain or of specific tissues. In this section, we present one of the possible applications of information geometry on manifold of probability distributions and demonstrate the use of probability distributions in the context of the classification of the Alzheimer's disease population.

5.1. Study set-up and design

Data used in the preparation of this paper were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database. ADNI is a project that has been initiated in 2004 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB) and the Food and Drug Administration (FDA), whose principal investigator is Dr. Michael Weiner. ADNI provides all data without embargo to all scientists in the world. The aim of the project is the development of clinical, genetic, biochemical or imaging biomarkers for the early diagnosis and follow-up of Alzheimer's disease. For up-to-date information, see ADNI website.

5.2. Participants

Our study is based on a part of ADNI population. Indeed, the initial subjects were not age- and sex-matched and our procedure consisted in randomly selecting subjects. In addition, some of the subjects were excluded because of a low diagnosis reliability (according to ADNI criteria) and others because of unsuccessful cortical thickness measurement due to poor image quality. The resulting population is composed of 143 subjects; 71 healthy controls (HC) subjects and 72 Alzheimer's disease (AD) patients whose demographic data are presented in Table 1.

5.3. MRI Acquisition

MRI volumes were downloaded from ADNI1 (i.e. ADNI first study). All the MR scans are T1-weighted MR images and were acquired on a 1.5 Tesla scanner. For each subject, we only used the MRI scan from the baseline visit and the ones that were acquired according to 3D MP-RAGE (Magnetization -Prepared Rapid Acquisition Gradient Echo) sequence. The 3D MP-RAGE sequence was used with the following protocol parameters: slice width = 1.2mm; echo time (TE)=3.61ms; repetition time (TR)=3000 ms; flip angle=8deg; matrix size=192x192; slice number=160-170; FOV=250mm; pixel size=1.25mm$\times$1.25mm. The MPRAGE images are considered the best in the quality ratings and have undergone gradwarping, intensity correction, and have been scaled for gradient drift using the phantom data.

5.4. Cortical thickness measurement and distribution

Cortical thickness was chosen as the MRI biomarker because of its ability to quantify morphological alterations of the cortical mantle in early stage of AD. Cortical Thickness (CTh) was measured using the Matlab Toolbox CorThiZon [16] and computed on the entire cortical ribbon using a Laplace's-equation-based algorithm as described by Jones *et al* [17]. Thus, a 3D cortical thickness map was obtained.

We applied the method of moments previously described to estimate the three generalized gamma parameters (α, λ, β) and thus we obtained the CTh distribution.

Table 1. Demographic and clinical characteristics of the study population

	HC (n=71)	AD (n=72)	p-value
Age (years)	76.1 ± 5.6	77.4 ± 5.5	0.17
Sex (F/M)	38 / 33	41 / 31	0.20
MMSE	29 ± 0.9	23.2 ± 2.1	<0.001

Plus-minus values are means $\pm$ standard deviation. All p-values are based on ANOVA test, apart from Sex, which is based on Chi-square tests (α <0.05). Abbreviations: HC, Healthy Control; AD, Alzheimer's disease patients; MMSE, Mini Mental State Examination.

5.5. Clustering Based on Distribution Similarity

Clustering, also called unsupervised classification, has been extensively studied for years in many fields, such as data mining, pattern recognition, image segmentation and bioinformatics. This technique is used primarily to segment or classify a database or extract knowledge to attempt to identify subsets of data that are difficult to distinguish. The aim is to group data sets in a way that the intra-cluster similarity is maximized while the inter-cluster similarity is minimized. Three principal categories of clustering exist in literature, namely partitioning clustering, hierarchical clustering and density-based clustering.

In our study, the experiments were conducted using partitioning k-medoids algorithm [18], that we extended using an approximate geodesic distance that is computed in two steps.
Let $p(\eta_2, \lambda_2, \beta_2), p(\eta_1, \lambda_1, \beta_1)$ be two generalized gamma densities. The energy E_1 of the path $t \in [0,1] \mapsto \gamma_\beta(t) = (\eta_1, \lambda_1, (1-t)\beta_1 + t\eta_2)$ is computed using the formula:

$$E_1 = (\beta_2 - \beta_1)^2 \int_0^1 g_{\beta\beta}(\gamma_\beta(t))dt$$

Then the energy E_2 of the path joining $p(\eta_1, \lambda_1, \beta_1)$ and $p(\eta_2, \lambda_2, \beta_2)$ is computed on the gamma submanifold only. The overall distance is then taken to be $\sqrt{E_1 + E_2}$. Using this approximate distance avoids circumvent numerical instabilities resulting from the positive curvature of the generalized gamma manifold in the plane $\partial_\eta, \partial_\lambda$ and yields a faster algorithm.

The K-medoids approach, as all clustering algorithm, tries to organize data into K clusters, to do so the method consists of two phases, the building phase and the swapping phase. The building phase consists on selecting the initial k representatives (i.e. medoids) at random. Non-selected objects are assigned to the most similar representative according to geodesic distance. Then, in the swapping phase, we iteratively replace representatives by non-representative objects (see algorithm 1).

Algorithm 1 Distribution based K-medoids algorithm

1. **Initialization:** Select randomly k distributions as the initial representative objects (i.e. k-medoids)
2. **Repeat**
   - *i.* Calculate the geodesic distance between each medoid m and the remaining data objects
   - *ii.* Assign the non representative object o_i to the closest medoid m (i.e. smallest geodesic distance)
   - *iii.* Compute the total cost S of swapping the medoid m with o_i; the total cost is defined to be the sum of the squared errors SSE of the resulting clustering
   - *iv.* If $S < 0$, then swap m with o_i to form the new set of medoids
3. **Until**
   Convergence criterion is satisfied (i.e. no change in the medoids or in total swapping cost)

The K-medoids algorithm is chosen instead of k-means algorithm for mainly two reasons: It minimizes a sum of pairwise dissimilarities instead of a sum of squared Euclidean distances. Consequently it is more robust to noise and outliers as compared to k-means. Moreover, k-means represent each cluster by the mean of all objects in this cluster, while k-medoids use an actual object in a cluster as its representative and since the objects in our case are probability distributions; it was more efficient to proceed with the k-medoids method [19].

5.6. Results

The quality of the clustering results was assessed using an external evaluation measure, called *Purity*. The external clustering measures are used to assess how well clusters matched up with real labels. In order to compute the evaluation measure *Purity*, each cluster is assigned to the class which is most frequent in the cluster, and then the accuracy of this assignment is measured by counting the number of correctly assigned objects and dividing by the total number of objects. It is the percent of the total number of objects that were classified correctly.

$$\text{Purity} = \frac{1}{N}\sum_{i=1}^{k}\max_{j}|c_i \cap t_j| \tag{22}$$

where N is the number of objects, k is the number of clusters, c_i is the number of objects in the i-th cluster of the clustering solution, and t_j the number of objects in the j-th cluster of the groundtruth c_i and $|c_i \cap t_j|$ is the number of objects in both the i-th cluster of the clustering solution and j-th cluster of the groundtruth. Figure1 summarizes the approach.

In our case, the aim was to assess how accurately our approach would group AD patients and HC subjects. Thus, we have chosen $k = 2$ as cluster number in the k-medoids algorithm, one cluster would represent the AD patients and the other the HC subjects. These clusters are compared with the true label data using the *Purity* measure. We obtained **Purity=0.84**, meaning that the two clusters of the distribution based k-medoids algorithm match up with 84% of the real labels.

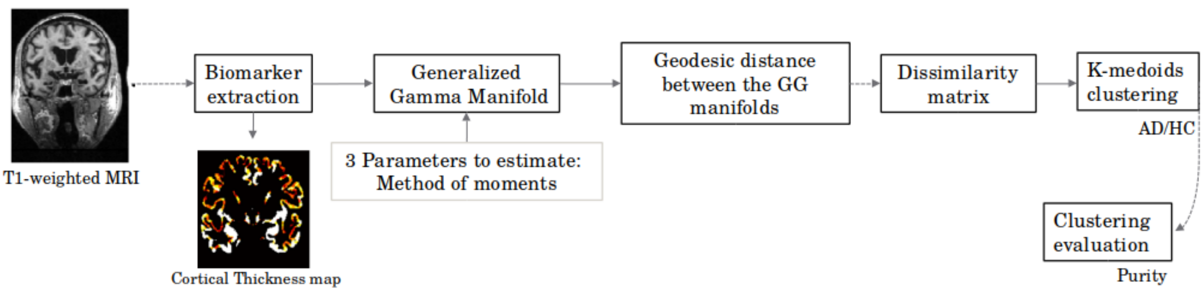

Figure 1. General scheme of the proposed approach

5.7. Discussion

The distribution based K-medoids algorithm obtains accurate classification of the Alzheimer's disease population (Purity = 84%). Indeed, information geometry offers suitable tools that allows the proper use of probability distribution, which increase significantly the performance of the disease classification compared to classical approaches [10]. Thus, we can consider that this approach is a powerful aid to study neurological diseases.

Author Contributions: F. Nicol and S. Puechmorel have contributed equally to the geometry of the generalized gamma manifold. S. Rebbah has written the part on medical applications.

Funding:

Acknowledgments: This work was supported by the Fondation pour la Recherche Medicale (FRM grant number ECO20160736068 to S.R).

Conflicts of Interest: 'The authors declare no conflict of interest

Abbreviations

The following abbreviations are used in this manuscript:

AD Alzheimer's Disease
CTh Cortical Thickness
HC Healthy Control
GG Generalized Gamma
MRI Magnetic Resonance Imaging

References

1. Stacy, E.W. A Generalization of the Gamma Distribution. The Annals of Mathematical Statistics **1962**, 33, 1187–1192.
2. Amoroso, L. Ricerche intorno alla curva dei redditi. Annali di Matematica Pura ed Applicata **1925**, 2, 123–159.
3. Crooks, G.E. The Amoroso Distribution. arXiv preprint arXiv:1005.3274 **2010**.
4. Shima, H. Geometry of Hessian Structures. Geometric Science of Information; Nielsen, F.; Barbaresco, F., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp. 37–55.
5. Duistermaat, J. On Hessian Riemannian structures. Asian journal of mathematics **2001**, 5, 79–91.
6. Arwini, K.; Dodson, C.; Doig, A.; Sampson, W.; Scharcanski, J.; Felipussi, S. Information Geometry: Near Randomness and Near Independence; Information Geometry: Near Randomness and Near Independence, Springer, 2008.
7. Chavel, I. Riemannian Geometry: A Modern Introduction; Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2006.
8. Guo, B.; F., Q.; J.L, Z.; Luo Q.M., journal=Mathematica Clovaca, v.y. Sharp Inequalities for Polygamma Functions.
9. Vemuri, P.; Jack, C.R. Role of structural MRI in Alzheimer's disease. Alzheimer's Research & Therapy **2010**, 2, 23. doi:10.1186/alzrt47.
10. Cuingnet, R.; Gerardin, E.; Tessieras, J.; Auzias, G.; Lehéricy, S.; Habert, M.O.; Chupin, M.; Benali, H.; Colliot, O.; Alzheimer's Disease Neuroimaging Initiative. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. NeuroImage **2011**, 56, 766–781. doi:10.1016/j.neuroimage.2010.06.013.
11. Lama, R.K.; Gwak, J.; Park, J.S.; Lee, S.W. Diagnosis of Alzheimer's Disease Based on Structural MRI Images Using a Regularized Extreme Learning Machine and PCA Features. Journal of Healthcare Engineering **2017**, 2017. doi:10.1155/2017/5485080.
12. Pini, L.; Pievani, M.; Bocchetta, M.; Altomare, D.; Bosco, P.; Cavedo, E.; Galluzzi, S.; Marizzoni, M.; Frisoni, G.B. Brain atrophy in Alzheimer's Disease and aging. Ageing Research Reviews **2016**, 30, 25–48. doi:10.1016/j.arr.2016.01.002.
13. Busovaca, E.; Zimmerman, M.E.; Meier, I.B.; Griffith, E.Y.; Grieve, S.M.; Korgaonkar, M.S.; Williams, L.M.; Brickman, A.M. Is the Alzheimer's disease cortical thickness signature a biological marker for memory? Brain imaging and behavior **2016**, 10, 517–523. doi:10.1007/s11682-015-9413-5.
14. Ruiz, E.; Ramírez, J.; Górriz, J.M.; Casillas, J.; Alzheimer's Disease Neuroimaging Initiative. Alzheimer's Disease Computer-Aided Diagnosis: Histogram-Based Analysis of Regional MRI Volumes for Feature Selection and Classification. Journal of Alzheimer's disease: JAD **2018**, 65, 819–842. doi:10.3233/JAD-170514.
15. Giulietti, G.; Torso, M.; Serra, L.; Spanò, B.; Marra, C.; Caltagirone, C.; Cercignani, M.; Bozzali, M.; Alzheimer's Disease Neuroimaging Initiative (ADNI). Whole brain white matter histogram analysis of diffusion tensor imaging data detects microstructural damage in mild cognitive impairment and alzheimer's disease patients. Journal of magnetic resonance imaging: JMRI **2018**. doi:10.1002/jmri.25947.
16. Querbes, O.; Aubry, F.; Pariente, J.; Lotterie, J.A.; Démonet, J.F.; Duret, V.; Puel, M.; Berry, I.; Fort, J.C.; Celsis, P.; Alzheimer's Disease Neuroimaging Initiative. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. Brain: A Journal of Neurology **2009**, 132, 2036–2047. doi:10.1093/brain/awp105.
17. Jones, S.E.; Buchbinder, B.R.; Aharon, I. Three-dimensional mapping of cortical thickness using Laplace's equation. Human Brain Mapping **2000**, 11, 12–32.

18. Kaufman, L.; Rousseeuw, P. Clustering by means of Medoids. Statistical Data Analysis Based on the L1–Norm and Related Methods, edited by Y. Dodge, North-Holland **1987**, pp. 405–416.
19. Soni, K.G.; Patel, D.A. Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data. International Journal of Computational Intelligence Research **2017**, 13, 899–906.

Sample Availability: Samples of the compounds ...... are available from the authors.